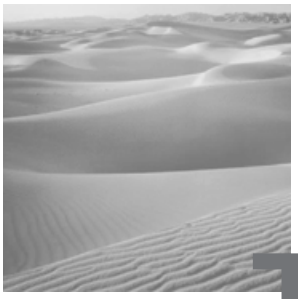


This second chapter will teach you the theory you must understand before continuing with the book. Appendix B will cover the Cisco Catalyst 1900 switch configuration, and Chapter 6 will cover Virtual LAN (VLAN) configuration. This chapter will give you the background you need to understand those chapters.

In this chapter you will learn the background behind the following topics:

- Layer-2 switching
- Address learning
- Forward/filtering decisions
- Loop avoidance
- Spanning-Tree Protocol
- LAN switch types

By reading and understanding the information presented in this chapter, you will be ready to configure switches and VLANs in Chapter 6 and Appendix B.

Layer-2 Switching

Layer-2 switching is hardware based, which means it uses the MAC address from the host's NIC cards to filter the network. Switches use Application-Specific Integrated Circuits (ASICs) to build and maintain filter tables. It is OK to think of a layer-2 switch as a multiport bridge. Layer-2

switches are fast because they do not look at the Network layer header information, looking instead at the frame's hardware addresses before deciding to either forward the frame or drop it.

Layer-2 switching provides the following:

- Hardware-based bridging (MAC)
- Wire speed
- Low latency
- Low cost

What makes layer-2 switching so efficient is that there is no modification to the data packet, only to the frame encapsulating the packet. Since no modification of the data packet is performed, the switching process is faster and less error-prone than routing.

Use layer-2 switching for workgroup connectivity and network segmentation (breaking up collision domains). This allows you to create a flatter network design with more network segments than traditional 10BaseT shared networks. Layer-2 switching increases bandwidth for each user because each connection (interface) into the switch is its own collision domain, so you can connect multiple devices to each interface.

Limitations of Layer-2 Switching

Since we think of layer-2 switching as the same as a bridged network, we must also think it has the same problems as a bridged network. Remember that bridges are good if we design the network correctly, meaning we break up the collision domains correctly. The right way to create bridged networks is to make sure that users spend 80 percent of their time on the local segment.

Bridged networks break up collision domains, but the network is still one large broadcast domain. Layer-2 switches (bridges) cannot break up broadcast domains, which can cause performance issues and limit the size of your network. Broadcasts and multicasts, along with the slow convergence of spanning tree, can cause major problems as the network grows. Because of these problems, layer-2 switches cannot completely replace routers (layer-3 devices) in the internetwork.

Bridging versus LAN Switching

Layer-2 switches are really just bridges with more ports. However, there are some important differences you should be aware of:

- Bridges are software based, while switches are hardware based because they use an ASICs chip to help make filtering decisions.
- Bridges can only have one spanning-tree instance per bridge, while switches can have many. (We cover spanning tree later in this chapter.)
- Bridges can only have up to 16 ports, whereas a switch can have hundreds.

Three Switch Functions at Layer 2

There are three distinct functions of layer-2 switching:

Address learning Layer-2 switches and bridges remember the source hardware address of each frame received on an interface and enter this information into a MAC database.

Forward/filter decisions When a frame is received on an interface, the switch looks at the destination hardware address and finds the exit interface in the MAC database.

Loop avoidance If multiple connections between switches are created for redundancy, network loops can occur. The Spanning-Tree Protocol (STP) is used to stop network loops and allow redundancy.

Address learning, forward and filtering decisions, and loop avoidance are discussed in detail in the next sections.

Address Learning

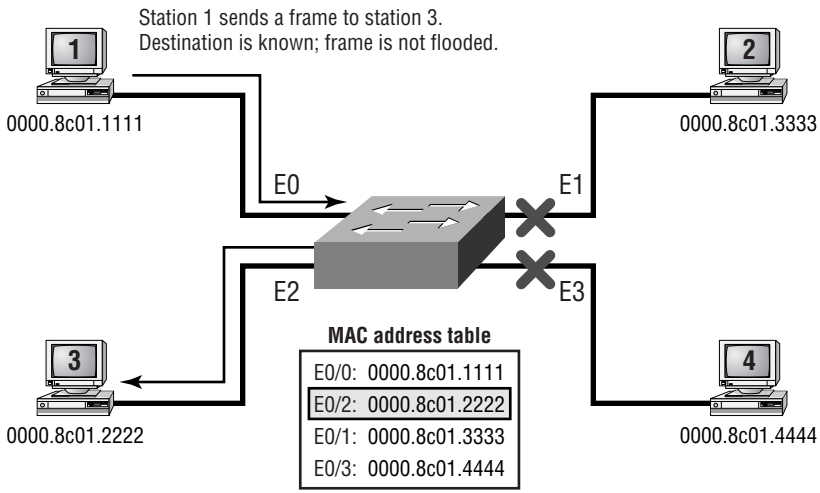
When a switch is powered on, the MAC filtering table is empty. When a device transmits and an interface receives a frame, the switch places the source address in the MAC filtering table, remembering what interface the device is located on. The switch has no choice but to flood the network with this frame because it has no idea where the destination device is located.

If a device answers and sends a frame back, then the switch will take the source address from that frame and place the MAC address in the database, associating this address with the interface that received the frame. Since the

switch now has two MAC addresses in the filtering table, the devices can make a point-to-point connection, and the frames will only be forwarded between the two devices. This is what makes layer-2 switches better than hubs. In a hub network, all frames are forwarded out all ports every time.

Figure 2.1 shows the procedures for how a MAC database is built.

FIGURE 2.1 How switches learn hosts' locations



In this figure, there are four hosts attached to a switch. When the switch is powered on, it has nothing in the MAC address table.

1. Host 1 sends a frame to Host 3. Host 1's MAC address is 0000.8c01.1111; Host 3's MAC address is 0000.8c01.2222.
2. The switch receives the frame on the E0/1 interface (interface addressing is covered in Appendix B) and places the source address in the MAC address table.
3. Since the destination address is not in the MAC database, the frame is forwarded out all interfaces.
4. Host 3 receives the frame and responds to Host 1. The switch receives this frame on interface E0/3 and places the source hardware address in the MAC database.

5. Host 1 and Host 3 can now make a point-to-point connection and only the two devices will receive the frames. Hosts 2 and 4 will not see the frames.

If the two devices do not communicate to the switch again within a certain amount of time, the switch will flush the entries from the database to keep it as current as possible.

Forward/Filter Decisions

When a frame arrives at a switch interface, the destination hardware address is compared to the forward/filter MAC database. If the destination hardware address is known and listed in the database, the frame is only sent out the correct exit interface. The switch does not transmit the frame out any interface except for the destination interface. This preserves bandwidth on the other network segments and is called *frame filtering*.

If the destination hardware address is not listed in the MAC database, then the frame is broadcasted out all active interfaces except the interface the frame was received on. If a device answers the broadcast, the MAC database is updated with the device location (interface).

Broadcast and Multicast Frames

Broadcast and multicast frames do not have a destination hardware address specified. The source address will always be the hardware address of the device transmitting the frame, and the destination address will either be all 1s (broadcast), or with the network or subnet address specified and the host address all 1s (multicast). For example, a broadcast and multicast in binary would be as shown in Table 2.1.

TABLE 2.1 Broadcast and Multicast Example

	Binary	Decimal
Broadcast	11111111.11111111.11111111.11111111	255.255.255.255
Multicast	10101100.00010000.11111111.11111111	172.16.255.255

Notice that the broadcast is all 1s, but the multicast is not. They are both a type of broadcast, except that a multicast just sends the frame to a certain network or subnet and all hosts within that network or subnet, and a broadcast of all 1s sends the frame to all networks and hosts.

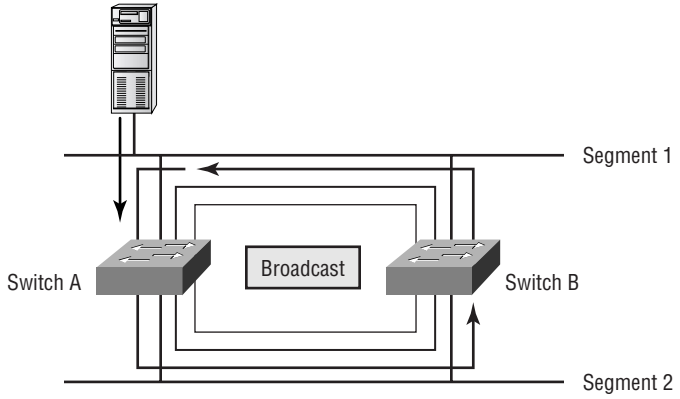
When a switch receives these types of frames, it is then quickly flooded out all of the switch's active ports by default. To have broadcasts and multicasts only forwarded out a limited amount of administratively assigned ports, you create Virtual LANs (VLANs), which are covered in Chapter 6.

Loop Avoidance

Redundant links are a good idea between switches. They are used to help stop complete network failures if one link fails. Even though redundant links are extremely helpful, they cause more problems than they solve. Because frames can be broadcast down all redundant links simultaneously, network loops can occur, among other problems. Some of the most serious problems are discussed in the following list.

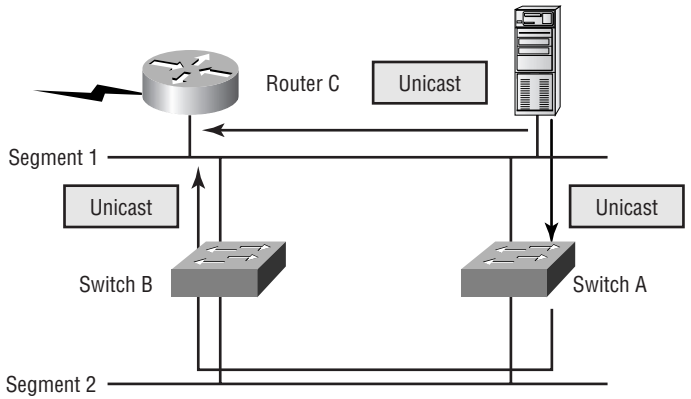
1. If no loop avoidance schemes are put in place, the switches will flood broadcasts endlessly throughout the internetwork. This is sometimes referred to as a *broadcast storm*. Figure 2.2 shows how a broadcast may be propagated throughout the network. Notice in the figure how a frame is continually broadcast through the internetwork Physical network.

FIGURE 2.2 Broadcast storms



2. A device can receive multiple copies of the same frame since the frame can arrive from different segments at the same time. Figure 2.3 shows how multiple frames can arrive from multiple segments simultaneously.

FIGURE 2.3 Multiple frame copies



3. The MAC address filter table will be confused about where a device is located since the switch can receive the frame from more than one link. It is possible that the switch can't forward a frame because it is constantly updating the MAC filter table with source hardware address locations. This is called *thrashing* the MAC table.

4. One of the biggest problems is multiple loops generating throughout an internetwork. This means that loops can occur within other loops. If a broadcast storm were to then occur, the network would not be able to perform packet switching.

The Spanning-Tree Protocol, discussed in the following section, was developed to solve the problems presented in this list.

Spanning-Tree Protocol (STP)

Digital Equipment Corporation (DEC), which was purchased and is now called Compaq, was the original creator of Spanning-Tree Protocol (STP). The IEEE created their own version of STP called 802.1d. All Cisco switches run the IEEE 802.1d version of STP, which is not compatible with the DEC version.

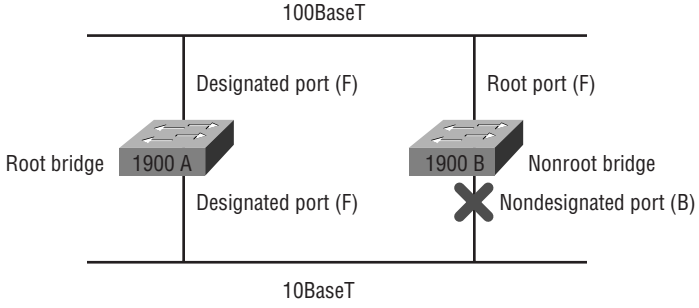
STP's main task is to stop network loops from occurring on your layer-2 network (bridges or switches). STP is constantly monitoring the network to find all links and make sure that loops do not occur by shutting down redundant links.

Spanning-Tree Operations

STP finds all links in the network and shuts down redundant links, thereby stopping any network loops from occurring in the network. The way it does this is by electing a root bridge that will decide on the network topology. There can only be one root bridge in any given network. Root-bridge ports are called *designated ports*, which operate in what are called forwarding-state ports. Forwarding-state ports send and receive traffic.

Other switches in your network are called nonroot bridges, as shown in Figure 2.4. However, the port with the lowest cost (as determined by a link's bandwidth) to the root bridge is called a root port and sends and receives traffic.

FIGURE 2.4 Spanning-tree operations



Ports determined to have the lowest-cost path to the root bridge are called designated ports. The other port or ports on the bridge are considered nondesignated and will not send or receive traffic, which is called blocking mode.

Selecting the Root Bridge

Switches or bridges running STP exchange information with what are called Bridge Protocol Data Units (BPDUs). BPDUs send configuration messages using multicast frames. The bridge ID of each device is sent to other devices using BPDUs.

The bridge ID is used to determine the root bridge in the network and to determine the root port. The bridge ID is 8 bytes long and includes the priority and the MAC address of the device. The priority on all devices running the IEEE STP version is 32,768.

To determine the root bridge, the priorities of the bridge and the MAC address are combined. If two switches or bridges have the same priority value, then the MAC address is used to determine which one has the lowest ID. For example, if two switches, which I'll name A and B, both use the default priority of 32,768, then the MAC address will be used. If switch A's MAC address is 0000.0c00.1111.1111 and switch B's MAC address is 0000.0c00.2222.2222, then switch A would become the root bridge.

The following network analyzer output shows a BPDU transmitted from a 1900 switch. BPDUs are sent out every two seconds by default. That may seem like a lot of overhead, but remember that this is only a layer-2 frame, with no layer-3 information in the packet.

From reading Chapter 1 you should be able to look at this frame and notice that it is an 802.2 frame, not only because it tells you so in the frame, but because it uses an 802.3-length field with a DSAP and an SSAP field in the LLC header.

```
   Flags:          0x80  802.3
   Status:         0x00
   Packet Length:64
   Timestamp:      19:33:18.726314 02/28/2000
802.3 Header
   Destination:    01:80:c2:00:00:00
   Source:         00:b0:64:75:6b:c3
   LLC Length:     38
```

```
802.2 Logical Link Control (LLC) Header
  Dest. SAP:     0x42  802.1 Bridge Spanning Tree
  Source SAP:    0x42  802.1 Bridge Spanning Tree
  Command:       0x03  Unnumbered Information
802.1 - Bridge Spanning Tree
  Protocol Identifier:  0
  Protocol Version ID:  0
  Message Type:         0  Configuration Message
  Flags:                %00000000
  Root Priority/ID:     0x8000  /  00:b0:64:75:6b:c0
  Cost Of Path To Root: 0x00000000  (0)
  Bridge Priority/ID:   0x8000  / 00:b0:64:75:6b:c0
  Port Priority/ID:     0x80  /  0x03
  Message Age:                        0/256 seconds
(exactly 0seconds)
  Maximum Age:                        5120/256 seconds
(exactly 20seconds)
  Hello Time:                        512/256 seconds
(exactly 2seconds)
  Forward Delay:                       3840/256 seconds
(exactly 15seconds)
Extra bytes (Padding):
  ........              00 00 00 00 00 00 00 00
Frame Check Sequence:  0x2e006400
```

Once you get to the actual BPDU data, notice the cost of path to root. It is zero because this switch is actually the root bridge. We discuss path costs more in the following section.

Selecting the Designated Port

To determine the port or ports that will be used to communicate with the root bridge, you must first figure out the path cost. The STP cost is an accumulated total path cost based on the bandwidth of the links. Table 2.2 shows the typical costs associated with the different Ethernet networks.

TABLE 2.2 Typical Costs of Different Ethernet Networks

Speed	New IEEE Cost	Original IEEE Cost
10Gbps	2	1
1Gbps	4	1
100Mbps	19	10
10Mbps	100	100

The IEEE 802.1d specification has recently been revised to handle the new higher-speed links. The 1900 switches use the original IEEE 802.1d specifications.

Spanning-Tree Port States

The ports on a bridge or switch running the STP can transition through four different states:

Blocking Won't forward frames; listens to BPDUs. All ports are in blocking state by default when the switch is powered up.

Listening Listens to BPDUs to make sure no loops occur on the network before passing data frames.

Learning Learns MAC addresses and builds a filter table but does not forward frames.

Forwarding Sends and receives all data on the bridged port.

Typically, switch ports are in either blocking or forwarding state. A forwarding port has been determined to have the lowest cost to the root bridge. However, if the network has a topology change because of a failed link or even if the administrator adds a new switch to the network, the ports on a switch will be in listening and learning state.

Blocking ports are used to prevent network loops. Once a switch determines the best path to the root bridge, then all other ports will be in blocking state. Blocked ports still receive BPDUs.

```
802.2 Logical Link Control (LLC) Header
  Dest. SAP:     0x42  802.1 Bridge Spanning Tree
  Source SAP:    0x42  802.1 Bridge Spanning Tree
  Command:       0x03  Unnumbered Information
802.1 - Bridge Spanning Tree
  Protocol Identifier:  0
  Protocol Version ID:  0
  Message Type:         0  Configuration Message
  Flags:                %00000000
  Root Priority/ID:     0x8000  /  00:b0:64:75:6b:c0
  Cost Of Path To Root: 0x00000000  (0)
  Bridge Priority/ID:   0x8000  / 00:b0:64:75:6b:c0
  Port Priority/ID:     0x80  /  0x03
  Message Age:                        0/256 seconds
(exactly 0seconds)
  Maximum Age:                        5120/256 seconds
(exactly 20seconds)
  Hello Time:                        512/256 seconds
(exactly 2seconds)
  Forward Delay:                       3840/256 seconds
(exactly 15seconds)
Extra bytes (Padding):
  ........              00 00 00 00 00 00 00 00
Frame Check Sequence:  0x2e006400
```

Once you get to the actual BPDU data, notice the cost of path to root. It is zero because this switch is actually the root bridge. We discuss path costs more in the following section.

Selecting the Designated Port

To determine the port or ports that will be used to communicate with the root bridge, you must first figure out the path cost. The STP cost is an accumulated total path cost based on the bandwidth of the links. Table 2.2 shows the typical costs associated with the different Ethernet networks.

TABLE 2.2 Typical Costs of Different Ethernet Networks

Speed	New IEEE Cost	Original IEEE Cost
10Gbps	2	1
1Gbps	4	1
100Mbps	19	10
10Mbps	100	100

The IEEE 802.1d specification has recently been revised to handle the new higher-speed links. The 1900 switches use the original IEEE 802.1d specifications.

Spanning-Tree Port States

The ports on a bridge or switch running the STP can transition through four different states:

Blocking Won't forward frames; listens to BPDUs. All ports are in blocking state by default when the switch is powered up.

Listening Listens to BPDUs to make sure no loops occur on the network before passing data frames.

Learning Learns MAC addresses and builds a filter table but does not forward frames.

Forwarding Sends and receives all data on the bridged port.

Typically, switch ports are in either blocking or forwarding state. A forwarding port has been determined to have the lowest cost to the root bridge. However, if the network has a topology change because of a failed link or even if the administrator adds a new switch to the network, the ports on a switch will be in listening and learning state.

Blocking ports are used to prevent network loops. Once a switch determines the best path to the root bridge, then all other ports will be in blocking state. Blocked ports still receive BPDUs.

If a switch determines that a blocked port should now be the designated port, it will go to listening state. It will check all BPDUs heard to make sure that it won't create a loop once the port goes to forwarding state.

Convergence

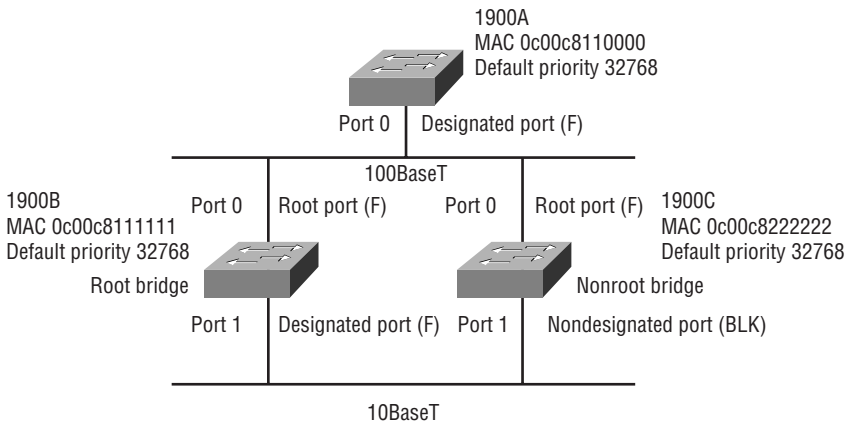
Convergence occurs when bridges and switches have transitioned to either the forwarding or blocking states. No data is forwarded during this time. Convergence is important to make sure all devices have the same database.

Before data can be forwarded, all devices must be updated. The problem with convergence is the time it takes for these devices to update. It usually takes 50 seconds to go from blocking to forwarding state. It is not recommended that you change the default STP timers, but the timers can be adjusted if necessary. Forward delay is the time it takes to transition a port from listening to learning state or from learning to forwarding state.

Spanning-Tree Example

It is important to see how spanning tree works in an internetwork, and this section will give you a chance to observe it in a live network. In Figure 2.5, the three switches all have the same priority of 32,768. However, notice the MAC address of each switch. By looking at the priority and MAC addresses of each switch, you should be able to determine the root bridge.

FIGURE 2.5 Spanning-tree example



Since 1900A has the lowest MAC address and all three switches use the default priority, then 1900A will be the root bridge.

To determine the root ports on switches 1900B and 1900C, you need to examine the cost of the link connecting the switches. Because the connection from both switches to the root switch is from port 0 using a 100Mbps link and has the best cost, both switches' root ports will be port 0.

To determine the designated ports on the switches, the bridge ID is used. The root bridge always has all ports as designated. However, since both 1900B and 1900C have the same cost to the root bridge, the designated port will be on switch 1900B since it has the lowest bridge ID. Because 1900B has been determined to have the designated port, switch 1900C will put port 1 in blocking state to stop any network loop from occurring.

LAN Switch Types

The latency for packet switching through the switch depends on the chosen switching mode. There are three switching modes:

Store and forward The complete data frame is received on the switch's buffer, a CRC is run, and then the destination address is looked up in the MAC filter table.

Cut-through The switch only waits for the destination hardware address to be received and then looks up the destination address in the MAC filter table.

FragmentFree The default for the Catalyst 1900 switch, it is sometimes referred to as modified cut-through. Checks the first 64 bytes of a frame for fragmentation (because of possible collisions) before forwarding the frame.

Figure 2.6 shows the different points where the switching mode takes place in the frame.

The different switching modes are discussed in detail in the following sections.